

ECON-SHU 310 Econometrics

Names: Cinny Lin (ycl461), Yihan Xu (yx1708), Yizhou Lu (yl5438)

October 4, 2020

1 Question 1

We throw a fair die with six sides twice independently. The faces of the die contain the numbers 1, 2, 3, 4, 5 and 6. Let X be the smallest number in these two throws and Y be the largest number.

1. Give the joint probability distribution of X and Y .

This table shows the probability of $P(X = x, Y = y)$:

$x \backslash y$	1	2	3	4	5	6	$P(x)$
1	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{11}{36}$
2	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{4}$
3	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{7}{36}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{5}{36}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$
$P(y)$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{1}{4}$	$\frac{11}{36}$	

From this table, we get that:

$$\mathbb{P}(X = x, Y = y) = \begin{cases} 0 & \text{for } x > y \\ \frac{1}{36} & \text{for } x = y \\ \frac{1}{18} & \text{for } x < y \end{cases}$$

2. Calculate the marginal probability distributions of X and Y .

The marginal probability distribution is the last column and row on the previous joint probability distribution table.

For example, $P(X = 1) = P(X = 1|Y = 1) + P(X = 1|Y = 2) + P(X = 1|Y = 3) + P(X = 1|Y = 4) + P(X = 1|Y = 5) + P(X = 1|Y = 6) = \frac{1}{36} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} = \frac{11}{36}$.

Therefore, we can get:

$$\begin{aligned} P(X = 1) &= \frac{11}{36}, P(X = 2) = \frac{1}{4}, P(X = 3) = \frac{7}{36}, \\ P(X = 4) &= \frac{5}{36}, P(X = 5) = \frac{1}{12}, P(X = 6) = \frac{1}{36}. \\ P(Y = 1) &= \frac{1}{36}, P(Y = 2) = \frac{1}{12}, P(Y = 3) = \frac{5}{36}, \\ P(Y = 4) &= \frac{7}{36}, P(Y = 5) = \frac{1}{4}, P(Y = 6) = \frac{11}{36}. \end{aligned}$$

3. Calculate the covariance of X and Y .

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$E(X) = \sum_{i=1}^6 X_i P(X_i) = \frac{11}{36} * 1 + \frac{1}{4} * 2 + \frac{7}{36} * 3 + \frac{5}{36} * 4 + \frac{1}{12} * 5 + \frac{1}{36} * 6 = \frac{91}{36}$$

$$E(Y) = \sum_{i=1}^6 Y_i P(Y_i) = \frac{1}{36} * 1 + \frac{1}{12} * 2 + \frac{5}{36} * 3 + \frac{7}{36} * 4 + \frac{1}{4} * 5 + \frac{11}{36} * 6 = \frac{161}{36}$$

$$\begin{aligned} E(XY) &= \sum_x xP(X = x)E(Y|X = x) = 1 * (1 * \frac{1}{36} + 2 * \frac{1}{18} + 3 * \frac{1}{18} + 4 * \frac{1}{18} + 5 * \frac{1}{18} + \\ &6 * \frac{1}{18}) + 2 * (2 * \frac{1}{36} + 3 * \frac{1}{18} + 4 * \frac{1}{18} + 5 * \frac{1}{18} + 6 * \frac{1}{18}) + \dots = \frac{49}{4} \end{aligned}$$

$$\text{Therefore, } Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{49}{4} - \frac{91}{36} * \frac{161}{36} = \frac{1225}{1296}.$$

4. Calculate $E(Y|X = 2)$.

$$\begin{aligned} E(Y|X = 2) &= \sum_y yP(X = 2, Y = y) = \sum_y y \frac{P(X=2, Y=y)}{P(X=2)} \\ &= \frac{1*0+2*\frac{1}{36}+3*\frac{1}{18}+4*\frac{1}{18}+5*\frac{1}{18}+6*\frac{1}{18}}{\frac{1}{4}} = \frac{38}{9} \end{aligned}$$

2 Question 2

On the packaging of coffee packages from Nestle, it states that the weight of the package is 500 grams. In practice, a pack of coffee will of course not weigh 500 grams precisely. Consider an experiment involving weight notes from a randomly selected pack of coffee. Assume that the outcomes space is equal to $S = [495; 505]$. The probability density function (*pdf*) is given by:

$$f(w) = \begin{cases} \frac{1}{25}(w - 495) & \text{for } 495 \leq w \leq 500 \\ \frac{1}{25}(505 - w) & \text{for } 500 < w \leq 505 \end{cases}$$

1. Show that $f(w)$ is indeed a probability density function.

(Hint: show $f(w) \geq 0$ for $\forall \omega$ and $\int f(w)dw = 1$.)

(a) $f(w) \geq 0$ for $\forall \omega$

- When $495 \leq w \leq 500$, $\frac{1}{25}(w - 495)$
- When $500 < w \leq 505$, $\frac{1}{25}(505 - w)$

Therefore, $f(w) \geq 0 \forall \omega \in [495, 505]$.

(b) $\int f(w)dw = 1$

- $\int_{495}^{500} \frac{1}{25}(w - 495)dw = \frac{1}{25}((\frac{500*500}{2} - 495 * 500) - (\frac{495*495}{2} - 495 * 495)) = \frac{1}{2}$
- $\int_{500}^{505} \frac{1}{25}(505 - w)dw = \frac{1}{25}((505 * 500 - \frac{505*505}{2}) - (505 * 505 - \frac{500*500}{2})) = \frac{1}{2}$

Therefore, $\int_{495}^{505} f(w)dw = \frac{1}{2} + \frac{1}{2} = 1$

By (a), (b), $f(w)$ is a probability density function.

2. Calculate $P(W \leq 497)$ and $P(W = 498)$.

(a) $P(W \leq 497) = \int_{495}^{497} \frac{1}{25}(w - 495) = 0.08$

(b) $P(W = 498) = \int_{498}^{498} \frac{1}{25}(w - 495) = 0$

(For continuous random variable, the probability of a single point is 0.)

3. Calculate $E(W)$.

$$E(w) = \int_{495}^{500} \frac{1}{25} w(w - 495) dw + \int_{500}^{505} \frac{1}{25} w(505 - w) dw = \frac{1505}{6} + \frac{1495}{6} = 500$$

3 Question 3

Suppose that X_1 and X_2 are discrete random variables with joint pdf of the form

$$f(x_1; x_2) = c(x_1 + x_2); x_1 = 0; 1; 2; x_2 = 0; 1; 2$$

and zero otherwise.

$x_1 \backslash x_2$	0	1	2	total
0	0	1	2	3
1	1	2	3	6
2	2	3	4	9
total	3	6	9	18

$x_1 \backslash x_2$	0	1	2	$P(x_2)$
0	0	$\frac{1}{18}$	$\frac{2}{18}$	$\frac{1}{6}$
1	$\frac{1}{18}$	$\frac{2}{18}$	$\frac{3}{18}$	$\frac{2}{6}$
2	$\frac{2}{18}$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{6}$
$P(x_1)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	

1. Find the constant c .

- $\sum_{x_1=0}^2 \sum_{x_2=0}^2 f(x_1, x_2) = 1$
- $\rightarrow c \sum_{x_1=0}^2 \sum_{x_2=0}^2 (x_1 + x_2) = 1$
- $\rightarrow c \sum_{x_1=0}^2 (3x_1 + 3) = 1$
- $\rightarrow 3c \sum_{x_1=0}^2 (x_1 + 1) = 1$
- $\rightarrow 3c(1 + 2 + 3) = 1$
- $\rightarrow 18c = 1$
- $\rightarrow c = \frac{1}{18}$

2. Find the marginal pdf's of X_1 and X_2 .

$$(a) f_{X_1}(x_1) = \sum_{x_2=0}^2 2f(x_1, x_2) = \sum_{x_2=0}^2 2\frac{1}{18}(x_1 + x_2) = \frac{1}{18}(3x_1 + 3) = \frac{x_1+1}{6}$$

$$(b) f_{X_2}(x_2) = \sum_{x_1=0}^2 2f(x_1, x_2) = \sum_{x_1=0}^2 2\frac{1}{18}(x_1 + x_2) = \frac{1}{18}(3x_2 + 3) = \frac{1+x_2}{6}$$

Therefore, $f_1(0) = \frac{1}{6}$, $f_1(1) = \frac{2}{6}$, $f_1(2) = \frac{3}{6}$; $f_2(0) = \frac{1}{6}$, $f_2(1) = \frac{2}{6}$, $f_2(2) = \frac{3}{6}$.

3. Are X_1 and X_2 independent? Why or why not?

If X_1, X_2 were independent, $P(X_1, X_2) = P(X_1)P(X_2)$.

But because $(\frac{x_1+1}{6})(\frac{1+x_2}{6}) \neq \frac{x_1+x_2}{18}$, therefore, X_1 and X_2 are not independent.

4. Find $E(X_1|X_2)$ and $E(X_1)$, check if $E(X_1) = E(X_1|X_2)$.

- $E(X_1) = \sum xP(x) = 0 * \frac{1}{6} + 1 * \frac{2}{6} + 2 * \frac{3}{6} = \frac{4}{3}$
- $E(X_1|X_2 = x_2) = 0 * \frac{f(0,x_2)}{f_2(x_2)} + 1 * \frac{f(1,x_2)}{f_2(x_2)} + 2 * \frac{f(2,x_2)}{f_2(x_2)} = \frac{5+3x_2}{3+3x_2}$
(by Question 3.2, replace $f_2(x_2)$ with $\frac{1+x_2}{6}$)

Because,

- $E(X_1|X_2 = 0) = \frac{5}{3} \neq \frac{4}{3} = E(X_1)$
- $E(X_1|X_2 = 1) = \frac{4}{3} = \frac{4}{3} = E(X_1)$
- $E(X_1|X_2 = 2) = \frac{11}{9} \neq \frac{4}{3} = E(X_1)$

Therefore, $E(X_1) \neq E(X_1|X_2)$.

4 Question 4

Let two data sets (x_i, y_i) and (x_i^*, y_i^*) be related by $x_i^* = c_1 + c_2x$ and $y_i^* = c_3 + c_4y$ for all $i = 1, \dots, n$, where c_1, \dots, c_4 are known constants. This means that the differences between the two data sets are the location and the scale of measurement. Such data transformations are often applied in economic studies. For example, y_i may be the total variable production costs in dollars of a firm in month i and y_i^* the total production costs in millions of dollars. Then c_3 are the total fixed costs (in millions of dollars) and $c_4 = 10^6$. Suppose we have the linear relation $y = a + bx + u$ between y and x .

1. Show that y^* and x^* also have a linear relationship of $y^* = a^* + b^*x^* + u$. Derive the relation between the least squares estimators of (a, b) for the original data and (a^*, b^*) for the transformed data.

We know that $y = a + bx + u$ and $y^* = a^* + b^*x^* + u$.

We also know that $x_i^* = c_1 + c_2x_i$ and $y_i^* = c_3 + c_4y_i$.

We can infer that $\bar{x}^* = c_1 + c_2\bar{x}$ and $\bar{y}^* = c_3 + c_4\bar{y}$, and that \bar{y} is on \hat{y} , and \bar{y}^* is on \hat{y}^* .

$$\hat{b}^* = \frac{\sum_{i=1}^n c_2(x_i - \bar{x})c_4(y_i - \bar{y})}{\sum_{i=1}^n (c_2)^2(x_i - \bar{x})^2} = \frac{c_4}{c_2}\hat{b}$$

$$\hat{a}^* = \bar{y}^* - \hat{b}^*\bar{x}^* = c_3 + c_4\bar{y} - \frac{c_4}{c_2}(c_1 + c_2\bar{x})\hat{b} = c_4\bar{y} - \hat{b}c_4\bar{x} + c_3 - \frac{c_1c_4}{c_2}\hat{b} = c_4\hat{a} + c_3 - \frac{c_1c_4}{c_2}\hat{b}$$

Because \hat{b}^* and \hat{a}^* have linear relations with \hat{b} and \hat{a} , thus, we prove that y^* and x^*

also have a linear relationship $y^* = a^* + b^*x^* + u$.

2. Show that $\hat{y}^* = c_3 + c_4\hat{y}$.

$$\begin{aligned}\hat{y}^* &= \hat{a}^* + \hat{b}^*\hat{x}^* \\ &= (c_4\hat{a} + c_3 - \frac{c_1c_4}{c_2}\hat{b}) + (\frac{c_4}{c_2}\hat{b})(c_1 + c_2\frac{\hat{y}-\hat{a}}{\hat{b}}) \\ &= c_4\hat{a} + c_3 - \frac{c_1c_4}{c_2}\hat{b} + \frac{c_1c_4}{c_2}\hat{b} + \frac{c_4}{c_2}\hat{b}\frac{c_2(\hat{y}-\hat{a})}{\hat{b}} \\ &= c_4\hat{a} + c_3 + c_4\hat{y} - c_4\hat{a} \\ &= c_3 + c_4\hat{y}\end{aligned}$$

3. Show that the goodness-of-fit R^2 is invariant with respect to this transformation, i.e., for the y equation and y^* equation, R^2 is the same.

$$\begin{aligned}R^2 &= \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \\ R^{2*} &= \frac{\sum (\hat{y}_i^* - \bar{y}^*)^2}{\sum (y_i^* - \bar{y}^*)^2} \\ &= \frac{\sum [(c_3 + c_4\hat{y}_i) - (c_3 + c_4\bar{y})]^2}{\sum [(c_3 + c_4\hat{y}_i) - (c_3 + c_4\bar{y})]^2} \\ &= \frac{\sum (c_4)^2 (\hat{y}_i - \bar{y})^2}{\sum (c_4)^2 (\hat{y}_i - \bar{y})^2} \\ &= \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (\hat{y}_i - \bar{y})^2} \\ &= R^2\end{aligned}$$

5 Question 5

Please take a screen shot of your output window with the answers and add it to your pdf file.

Consider the data-set NLS-Y (US National Longitudinal Survey of Youth), containing a set of young men surveyed in 1980. In the dataset wage.dat, data are provided on RNS, MRT, SMSA, MED, KWW, IQ, AGE, S, EXPR, TENURE, LW. The definition of these variables is:

RNS: dummy for residency in the southern states

MRT: dummy for marital status

SMSA: dummy for residency in the metropolitan areas

MED: mother's education in years

KWW: score on the 'knowledge of the world of work' test

IQ: IQ score

AGE: age of the individual

S: completed years of schooling

EXPR: experience in years

TENURE: tenure in years

wage: monthly wage

1. What is the average and the median wage of the individuals in the sample? What is the average and the median log wage of the individuals in the sample?

hw1_q5.1.png

2. What is the median wage of the individuals for the subsample of individuals who live in metropolitan areas? please use two different approach to calculate this result.

hw1_q5.2.png

3. What is the covariance of IQ and S? Answer this question again for the subsample of individuals whose net income is lower than the 90th. percentile value of net income.

hw1_q5.3.png